View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

How to incorporate the spatial dimension in destination choice models? The case of Antwerpen

Hakim Hammadou[1], Isabelle Thomas[2] - University of Louvain-La-Neuve
Hans Tindemans[3], Frank Witlox[4] - Ghent University
Dries Van Hofstraeten[5], Ann Verhetsel[6] - University of Antwerp

Draft version – April 30th 2004
This is a work in progress. For latest update, please contact Hans Tindemans[3]

Abstract
This paper considers different alternatives for including spatial aspects within the activity-based approach for modeling destination choices. The study area is the urban agglomeration of Antwerpen (Belgium); the city and its suburbs are considered. Individual travel surveys are used. The paper pays particular attention to the inclusion of space within the decision context by including specific land-use explanatory variables generated by Geographical Information Systems. A preliminary geographical analysis is performed in order to represent the city by a limited set of destinations (*n* = 33) and to characterize those zones in terms of land use. Discrete choice modelling is used: each individual faces the total set of spatial destination alternatives. Several modelling approaches are explored and compared in terms of utility function (for instance Box-Cox; random coefficients) and in terms of global formulation (multinomial logit versus nested logit). The mixed nested logit formulation is selected as best and the parameter estimations are interpreted; it shows the importance of space within destination choices. This paper provides a useful background for decision-makers and planners of transportation policy related to individual mobility patterns.

Keywords
Discrete choice model, activity-based approach, GIS, land use, urban mobility, Antwerpen

1. Introduction

In activity-based modeling, researchers view travel as a derived demand for personal activities that are distributed in space and time (Jones et al., 1990; Axhausen and Garling, 1992). Travel decisions can also be considered as being part of a broader activity scheduling process. It is known that this approach leads to a better understanding of travel behavior compared to traditional modeling. It also enables a better analysis of possible responses to policies and their effect on traffic and spatial planning in urban development. Travel demand analysis is intrinsically spatial, i.e. spatial separation is the essence of travel demand. Therefore, in travel analysis and modeling, the spatial distribution of travel has recently been taken into account explicitly (Bates, 2000, p. 14). However, up to now, in most studies temporal aspects (e.g. the distribution of activities in time) have only been considered (Cirillo et al., 2002), whereas the spatial dimension clearly remains unexamined

[1] hammadou@geog.ucl.ac.be, Address : Place Louis Pasteur 3, 1348 Louvain-La-Neuve
[2] isabelle@geog.ucl.ac.be, Address : Place Louis Pasteur 3, 1348 Louvain-La-Neuve (corresponding author)
[3] hans.tindemans@ugent.be, Address : De Sterre Building S8, Krijgslaan 281, 9000 Gent
[4] frank.witlox@ugent.be, Address : De Sterre Building S8, Krijgslaan 281, 9000 Gent
[5] dries.vanhofstraeten@ua.ac.be, Address : Prinsstraat 13, 2000 Antwerpen
[6] ann.verhetsel@ua.ac.be, Address : Prinsstraat 13, 2000 Antwerpen

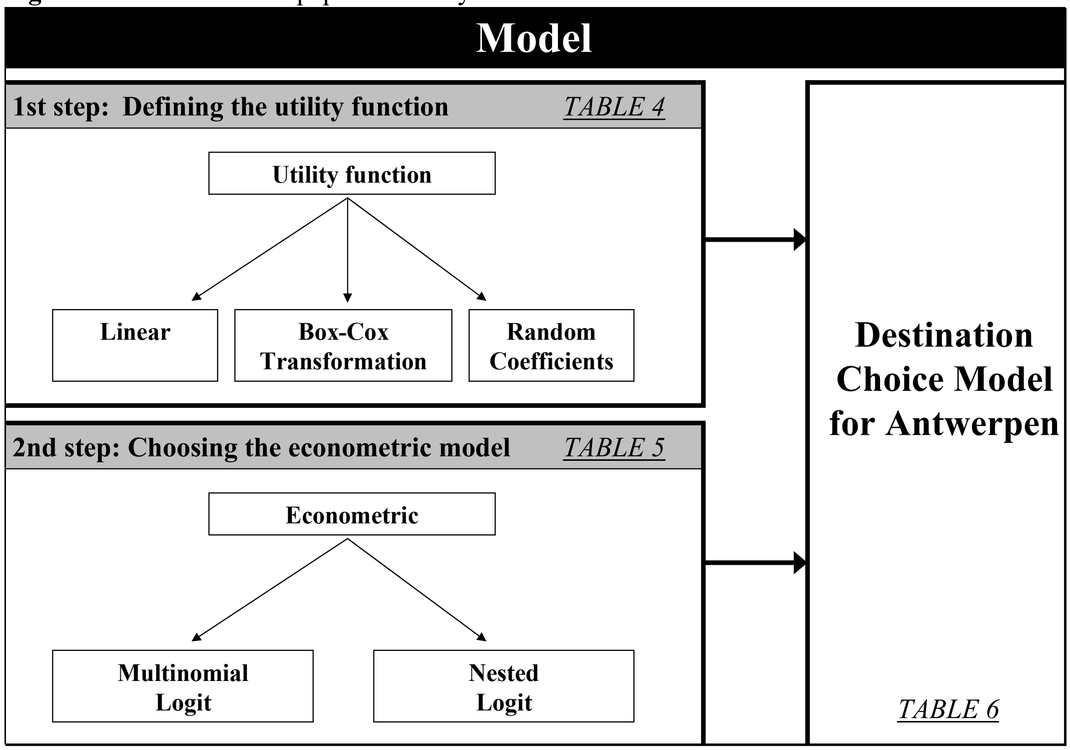
(Bhat and Zhao, 2002). We even ascertain that there is a lack of papers on activity chains models including spatial components. To our knowledge only a few exceptions exist: Bhat & Zhao (2002), McNally (2000) and Dijst & Vidakovic (1997). They consider the 'spatialisation' of the activities, but do not try to completely incorporate the spatial dimension in the model building process**.** Dijst and Vidakovic (1997) for example focus on people's action space in the city using only the spatial variable 'distance between locations of activity bases'. By explicitly adding the spatial dimension to travel resulting from household activities, the present research aims at making a valuable contribution to the modeling and understanding of travel behavior. Moreover, the objective is to obtain a complete spatially detailed and disaggregated destination choice model for the city region of Antwerpen. This research uses the OVG data set of the Antwerpen city region (1999) based on individual surveys describing the daily activities.

An important problem in destination choice modeling is the large number of choice alternatives. Ben-Akiva (1985) suggested to use a restricted set of alternatives rather than a full set. In this paper, we prefer to define a new zoning instead of using a sampling method. Therefore, the study area was divided into destination choice zones representing the alternative destinations. Instead of extracting the alternatives by using a random sampling approach, the total set of spatial destination alternatives for each individual was considered and aggregated into zones. We assume that the choice is dependent on the characteristics of the alternatives and the travelers. Hence, it is necessary to have information on attributes of the zones, and on individual and household characteristics. New data on the characteristics of land use, density and accessibility were generated using GIS techniques in order to have detailed information on the spatial structure of the study area (see Section 3).

The objective of a destination choice model is to give insights into the respondent's probability for choosing within a set of destination zones. The purpose of this study is to incorporate the real value of space in destination choice models, and hence to develop an appropriate modelling procedure. In order to obtain the best model fit, several sensitivity analyses are performed (see Section 4). First, the utility function is considered. We particularly consider (1) the statistical significance of the Box-Cox transformation compared to the linear formulation, and (2) a formulation of the model with random coefficients. The latter enables us to take into account the existence of heterogeneous preferences. Second, the performance of the multinomial logit is compared to that of the nested logit (Figure 1)

The paper is organized as follows. Section 2 discusses the methodological problems and defines choice models used. Section 3 is dedicated to the data set and the introduction of the spatial variables on land use, density and accessibility. Results of the destination choice models are presented in Section 4. Simulations are conducted to test the influence of spatial planning on destination choice. Conclusions and research perspectives are reported in Section 5.

Figure 1: Structure of the paper and analysis



2. Methodology

2.1. Model specification and estimation procedure

Discrete choice models assume that the global utility of a choice alternative is composed of a fixed (i.e. systematic or deterministic) utility value and a random or error utility component. Depending on the assumptions made regarding the distributions of the error terms, several discrete choice models have been developed in the literature (see e.g. Timmermans and Golledge, 1990).

The best known and also most applied discrete choice model is the MNL, the **multinomial logit model** (see Domencich and McFadden, 1975). In this model, the random utility elements are assumed to be (i) independently, (ii) identically and (iii) Type I extreme value (or Gumbel) distributed (McFadden, 1973, 1976). Independently and identically distributed (IID) error terms imply that the variances of the random components of the utilities are equal (homoscedasticity) and that all co-variances (or cross-effects) are assumed to be equal to zero. If IID can be defended, the Type I extreme value distribution seems the most suitable distribution (Johnson and Kotz, 1970). A detailed description can be found in Ben-Akiva and Lerman (1985), Ortuzar and Willumsen (2001); the basic ideas were developed by McFadden (1973).

The **conditional MNL** model for the destination choice analysis in the Antwerpen city region may be derived as follows. Assume that the utility that an individual *i* (= 1,..., I) derives from a choice alternative *j* (= 1, ..., J) is equal to:

$$U_{ij} = \beta_{jk} \boldsymbol{X}_{ijk} + \varepsilon_{ij} \quad [1]$$

where β_{jk} is the parameter for attribute *k* of alternative *j*, $\boldsymbol{X}_{ijk}$ is a vector of observable attribute values, and ε_{ij} represents the random unobserved component of utility.

Individual *i* will choose alternative *j* if the expected utility, U_{ij}, exceeds the expected utility, U_{im}, of alternatives *m*, where *m* indexes the elements of the set of alternatives. Hensher and Johnson (1981), Maddala (1983), Ben-Akiva and Lerman (1985), Train (1986) and Cramer (1991) have shown that, if the errors terms in equation [1] are IID according to a Type I extreme value distribution, the probability that individual *i* will choose alternative *j*, $P(j \mid C_i)$, is given by:

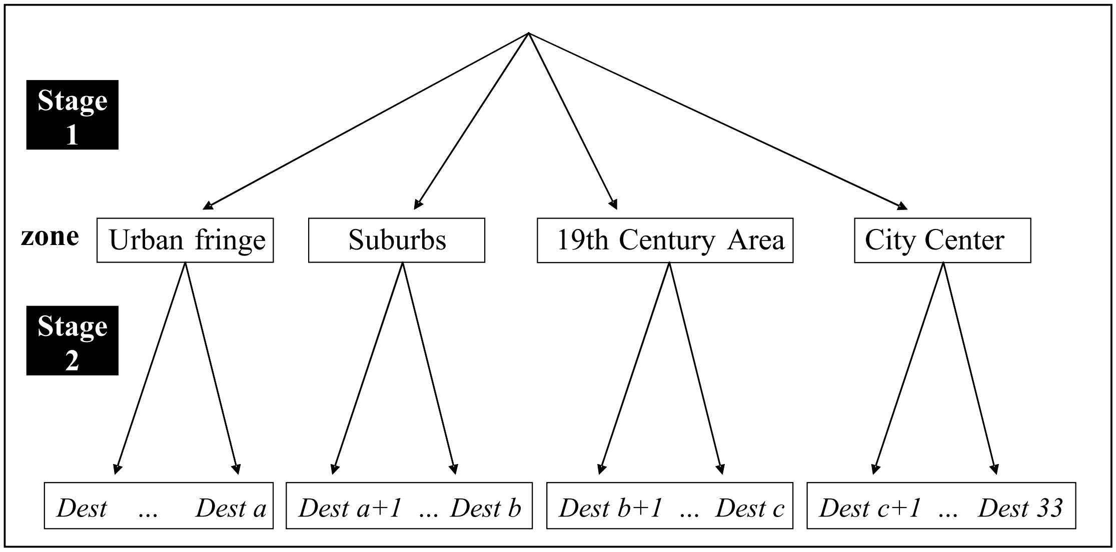
$$P(j \mid C_i) = \frac{\exp(\beta_{jk} X_{ijk})}{\sum_{m \in C_i} \exp(\beta_{mk} X_{imk})} \qquad [2]$$

Equation [2] can be estimated using a Maximum Likelihood method. It has been shown that this method produces consistent estimates of the parameters of the utility function as long as the disturbance terms are independent across all alternatives (McFadden, 1978).

As mentioned by Papola (2004), simple covariance matrix assumptions (e.g. homoscedastic diagonal matrix) gave rise to simple closed form models such as the MNL, while general covariance matrix assumptions have given rise to non-closed form models such as (among others) the **mixed multinomial logit model** (MXMNL) (Train and McFadden, 2000). MNL models have a strong theoretical base, have a simple mathematical structure and are also quite easy to estimate. However, they are based on restrictive hypotheses among which the IID which means that the model can only be applied to situations in which alternatives from which you can choose are totally independent. This is for sure not the case for spatial alternatives (our concern). Literature suggests the nested logit (NL) as a modelling alternative: choice alternatives are segmented, structured in branches that are more similar. Indeed, if we view the destination choice process as hierarchical and group similar alternatives into the same branches of the choice hierarchy, then establishments within each branch are more likely to follow the IID. However, at some exceptions, the mixed logit is rarely used in destination choice models (see e.g. Suarez et al, 2004).

In our case, we group spatial choice alternatives into subgroups. Each individual is supposed to first choose an urban level (e.g. urban versus suburban) and then, within that broad spatial zone, to choose a more precise destination. Hence, spatial choices are hierarchically organised (Figure 2). That is why in this paper the performance of the mixed logit will be compared to that of the multinomial logit.

Figure 2: Decision-making structure according to the proposed nested logit model of destination choices in Antwerp (*a*, *b*, *c* : number of destinations within each zone)



In order to interpret the nested logit model as a discrete choice model, we consider a random utility function. We assume that the utility function of the destination choice j can be split into a part that characterises the urban level and that does not vary with the choice within that level ($V_{j(l)}$)

$$U_j = V_l + V_{j(l)} + \varepsilon_j \qquad [3]$$

where ε_j is the stochastic part of the utility (error term). We assume that ε_1 , ε_2 ,…, ε_j (the individual specific error terms) are random and IID distributed following this distribution

$$F(\varepsilon_1, \varepsilon_2, ..., \varepsilon_J) = \exp\left(\sum_{l=1}^{L} \left(\sum_{j=1}^{J_l} e^{-\varepsilon_j / \lambda_l} \right)^{\lambda_l} \right) \qquad [4]$$

McFadden (1978) has shown that this formulation enables to write the model as a utility maximisation. In our case, the probability that an individual chooses a destination j is hence given by

$$P(j) = \frac{\exp\left(V_j / \lambda_l\right) \left(\sum_{j=1}^{J_l} \exp\left(V_{j(l)} / \lambda_l\right) \right)^{\lambda_l - 1}}{\sum_{l=1}^{L} \left(\sum_{j=1}^{J_l} \exp\left(V_{j(l)} / \lambda_l\right) \right)} \qquad [5]$$

which corresponds to the nested logit formulation with two levels of decision for destination.

Following Hensher and Greene (2002), the nested logit is indeed appealing in terms of its ability to accommodate differential degrees of interdependence (i.e. similarity) between subsets of alternatives in a choice set, but most published applications display a frequent lack of attention to the very precise form that these models must take to ensure that the resulting model is consistent with utility maximisation. As we do not know the utilities of the choice alternatives when building our destination choice model, two alternatives for the

utility functions were considered in this paper: the Box-Cox transformation and the random coefficients.

The search for a utility function is restricted to travel time. This choice was made a priori and it was recently and totally independently confirmed by Suarez *et al.* (2004). Let us here remind that the Box-Cox transformation of a variable x can be written as

$$\begin{aligned} BC_{\lambda}(x) &= \frac{x^{\lambda}-1}{\lambda} \text{ and } \lambda \neq 0, x>0 \\ BC_{\lambda}(x) &= \ln(x) \text{ and } \lambda = 0, x>0 \end{aligned} \qquad [6]$$

Using this transformation, one alleviates the restriction of linearity.

In a second set of sensitivity analyses, we consider random coefficients. Indeed, in a MNL formulation coefficients are constant (fixed for all individuals): the explanatory variables have the same effect for each individual. However, the population is by definition heterogeneous and the effect of each explanatory variable can vary from one individual to another. Let us give the example of travel time: it can depend upon non observable characteristics of the individuals (Revelt and Train, 1999). Hence the utility function associated to a destination j for an individual i can be considered.

$$U_{ij} = \gamma T_{ij} + X_{ij}\beta + \varepsilon_{ij} \qquad [7]$$

For each individual i and for each destination j, T_{ij} is the travel time, X_{ij} is a vector of observable attributes and β is the vector of parameters to be estimated. We assume that residuals ε_{ij} follow a Gumbel distribution. γ is the parameter which measures the travel time effect; we further assume that it is random and follows a normal distribution. The density function *f(y)* can be written as follows :

$$f(\gamma) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\gamma-\omega}{\sigma}\right)^2\right]}{\sigma\sqrt{2\pi}} = n(\omega,\sigma^2) \qquad [8]$$

where $n(\omega,\sigma^2)$ is the density function of a normal distribution with mean ω and variance σ^2.

The conditional probability to choose a destination j will be given by :

$$P_n(j \mid \gamma) = \frac{\exp(\gamma T_{jn} + X_{jn}\beta)}{\sum_{j=1}^{J}\exp(\gamma T_{jn} + X_{jn}\beta)} \qquad [9]$$

The unconditional probability to choose a destination j is obtained by integration of γ :

$$P_n(j) = \int_0^{\infty} P(j \mid \gamma).n(\omega,\sigma^2)d\gamma \qquad [10]$$

However, since the error term follows a Gumbel distribution, this model is based on the hypothesis of IID. This model is also called mixed logit with random coefficients.

2.2. Activity and tour structure and the intermediate stop model

A first step in the estimation of disaggregated destination choice model consists in defining the choice set for activity patterns. There are numerous possible activity structures and tour structures that a person can make. One of the many decisions in each activity-based model is how to simplify and aggregate the various structures to a reasonable (limited) number of choice alternatives. In our model it was decided to consider only the destination from the main stop in the tour. Work is a mandatory activity and often fixed in space. Due to this deterministic character of work, only non-work stops are here considered. It is not possible to apply probabilistic models to explain this type of mandatory activities.

The four possible tour structures are:

- Home-Main-Home
- Home-Intermediate-Main-Home
- Home-Main-Intermediate-Home
- Home-Intermediate-Main-Intermediate-Home

The model structure considers the choice of the main destination of the tour. The alternatives in the destination choice models determine the tour structure and stops for various purposes. Our model estimates the probability that a person who makes a stop in a tour chooses a specific zone as his/her destination. The model includes stops made before and after the main activity in the tour. Table 1 shows the distribution of these four types of tour structures in the studied data set.

Table 1: Distribution of tour types

Tour type	Frequency	Percentage
Home-Main-Home	2889	80,0%
Home-Intermediate-Main-Home	359	9,9%
Home-Main-Intermediate-Home	246	6,8%
Home-Intermediate-Main-Intermediate-Home	115	3,3%
Total	3609	100,0%

Source: OVG Antwerpen city region (1999)

3. Data sets and definition of destination zones

In our model, a key role is assigned to the spatial characteristic of the studied area. To that end, spatial variables are incorporated into the model in two ways. First, the definition of the choice alternatives is based on the spatial characteristics of the study area. Spatial data generated from land use, density and accessibility data sets will be analyzed. The study area is divided into 33 homogeneous destination choice zones. Second, the spatial characteristics of destinations zones are introduced in the model as explanatory variables.
In this section, we first define the travel data set (Section 3.1), then the study area and the spatial variables (Section 3.2), and lastly, the destination choice zones using the spatial information (Section 3.3).

3.1. Travel data

The travel data set used in this research is the so-called OVG travel data set of the Antwerpen city region collected in 1999 (part of the Flemish Travel Behavior Research-project). In this survey, each person above the age of five and being a member of the selected sample of households is asked to fill in a travel diary for two consecutive days. This resulted in a large data set, including data on each trip (e.g. activity, mode, distance, duration) as well as socio-demographic information on each person and household (e.g. age, income, household type, sex). In sum, the Antwerpen data set contains information of about 30,000 trips made by 5,613 different persons (Verhetsel *et al.*, 2004).

3.2. Study area and spatial variables

Our study area is defined by looking at the location of the trip destinations of the respondents of this survey. It consists in 608 statistical sectors (also called neighborhoods or wards) that are equivalent to 32 zip codes or 18 communes. On the sector level, several spatial characteristics were generated: variables on land-use, density and accessibility. The characteristics of space, the density or availability of functions and services and the accessibility of functions and facilities in other areas have also an impact on travel behavior and destination choice. Van Wee (2002) stressed the importance of the introduction of spatial variables in the analyses of travel behavior. Badoe and Miller (2000) and Stead (2001) also indicated that land use, density and accessibility are three important groups of spatial variables that explain travel choice behavior. After generating the spatial attributes, the 608 sectors are aggregated to a limited set of homogeneous destination choice zones by using a spatial zoning algorithm which is explained hereafter.

3.2.1 Land-use variables

The first category of spatial variables is land use. In 1996 the OC-GIS Flanders developed a first digital land use map for the Flemish region in the framework of the federal research program TELSAT. In 2001, this land use data set has been updated in order to analyze changes in land use. The data set is based on satellite images, soil information and the road network. By using an automatic classification procedure, satellite information is converted into 19 categories of land use (see Table 2*)* (see OC-GIS Vlaanderen, 2002). A second data set concerning land use was used: the MultiNet data set (2001) collected by TeleAtlas which consists of administrative borders, the road network and specific land use information such as "built-up area".

Let us now compute for each of the 608 statistical sectors the surface occupied by each type of land use (in square meters as well as in percentage of the total surface). Each sector receives the same items of land use that can be compared with each other and with percentages of other sectors. The most interesting variables are (i) housing development (densely, built-up, green residential,…), (ii) industrial, commercial and port development, (iii) green areas and open spaces**,** and (iv) infrastructure (highways, airports, railways,…).

Table 2: Overview of the 19 categories of land use

1	**Agriculture/Meadowland**	Agriculture and open space
2		Meadowland
3		Alluvial meadowland
4		Orchards
5	**Forests and parks**	Coniferous forests
6		Broad-leaved forests
7		Mixed forests
8		Municipal parks
9	**Development and industry**	Densely-built housing
10		Housing and other development
11		Industrial and commercial area
12	**Infrastructure**	Highways
13		District roads
14		Airport infrastructure
15		Port infrastructure
16		Other infrastructure (railways,…)
17	**Heathland and dunes**	Heathland
18		Dunes
19	**Water**	

Source: Land cover and land use data set - OC-GIS VLAANDEREN (2002)

3.2.2 Density variables

A second group of spatial variables consists in density variables, which give an indication of the population density, employment, shopping and schools. Given that the shape and size of the statistical sectors are not identical - cf. MAUP, (Openshaw and Taylor, 1979) - density is here roughly defined as the number of people, jobs, dwellings,… per square kilometer. The expected effect on travel behavior is clear: the higher the density, the more destinations within the activity range, thus, saving time and money of the respondent. In other words, a high density enables people to participate in more activities during a giving time range (Van Wee, 2002).

Cervero (1996), Cervero and Knockelman (1997) and Badoe and Miller (2000) also found that a high **population density** leads to shorter trip distances and the discouragement of the use and possession of cars; they discussed the relationship between population density and travel behavior. That is why it is interesting to attach the population density to each origin and each destination sector. The data used here are those provided by the National Institute of Statistics (NIS, 2001).

In the literature the link between **employment**, travel behavior and activity patterns also received some attention. Badoe and Miller (2000, p. 251) indicated a consistency in the research results: a higher concentration of employment has a significant impact on travel behavior. The data set used to compute employment density is the employment data set of the Regional Development Agency of 2001. This is only a rough but workable estimation of the exact employment figures of the National Census of 2001.

A third density variable is the **school density** for each sector in the city region. The data set of the Department of Education of the Flemish Government contains a list of addresses for all the Flemish schools (primary and secondary schools, colleges, universities etc.). After geo-coding the addresses, it was possible to compute the number of schools per square kilometer. A high density of schools will no doubt have an impact on the modal choice (i.e. bringing or getting children from school) and the distance of school trips. In future research, data on the size of schools (number of pupils) will be taken into account to calculate school density variables.

A final density variable is **shopping density**. This is more difficult to introduce since complete data sets on commercial activities are not easy to obtain. However this is a crucial variable in our analysis since shopping trips frequently appear in our travel data set. We expect high shopping density to lead to shorter trips and lower use of cars. An internet data source (SCOOT) gave us the opportunity to find addresses and extra information on shopping alternatives. This resulted in approximately 6,000 stores (great and small) in the Antwerpen city region. Geo-coding these addresses made it possible to assign this information to the sectors and to compute shopping density. Data on the size of shops (store surface) will be available in the near future.

3.2.3 Accessibility variables

Between each centroid of the 608 sectors the shortest path network distance and the fastest path network time were generated on the basis of the *StreetNet* 2001 network and by using *ArcView Network Analyst*. The *StreetNet* road network includes information on traffic regulations, such as closed streets, one-way streets, underpass and overpass and travel cost. We did however not account for congestion, waiting time at traffic lights or extra time to take turns. To partly compensate, slightly lower than the actual maximum speed were assigned to all streets. To generate the shortest route between each centroid, *ArcView Network Analyst* implements a modified Dijkstra algorithm (ESRI, 2003; Sherlock et al., 2002; Wise, 2002). The *StreetNet* software enables one to compute two variables, namely network distance and travel time by car. For the other transport modes (foot and bike) network distance is used to calculate travel time. The assumptions made on speed factors are 4 km/h for walking and 15 km/h for biking.

The business service schedule of the local public company (De Lijn) was used to estimate the cost factor for public transport. The cost is calculated as the total time between two centroids of sectors in the study area, including walking time, weighting time and in-vehicle time. A second variable focuses on the availability and frequency of public transport. Using the frequency tables of De Lijn, i.e. a network of public transport lines and bus and/or tram stops, it was possible to compute for each statistical sector the frequency (either per day or per hour) of public transport lines at bus and/or tram stops located in the sector.

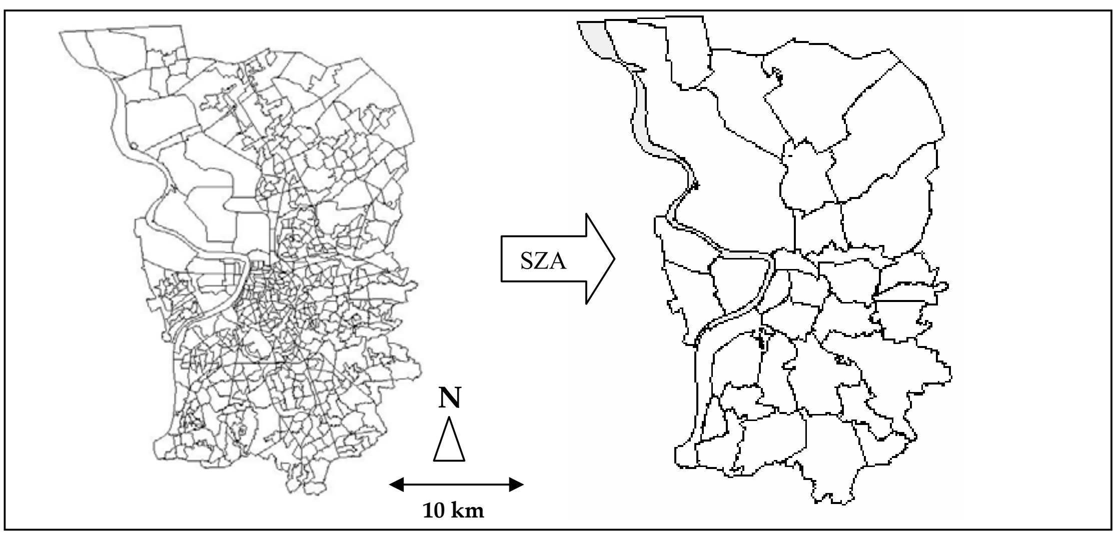
3.3. Defining destination choice zones

Travel demand models typically investigate the interactions among aggregated travel analysis zones within a study area. However, spatial aggregation can substantially affect the resulting travel demand modeling results. Miller (2004) indicated some major spatial analytical issues related to zoning: spatial dependency, spatial heterogeneity, boundary problems and scale effects. Unfortunately, there is no predefined method for zoning which avoids these analytical pitfalls. Hence, in the present paper, we attempt to tackle these fundamental spatial problems by constructing a zoning algorithm that aims at maximizing internal homogeneity and external heterogeneity of the aggregated zones. For computational tractability and data availability reasons, the study area is aggregated into 33 geographic analysis zones, based on GIS generated data.

The basic spatial unit is the statistical sector (or neighbourhood). Working at this level might cause problems due to the high number of sectors in the study area (*N*=608) and to their often small size. The level of the zip codes (mostly administrative borders) is also not suitable since zip code borders do not always take into account the characteristics of the spatial structure or the functional offer. That is why in this research the use of a Spatial Zoning Algorithm is advocated, i.e. an algorithm that is based on the results of a clustering or classification of sectors by their spatial characteristics as calculated in the previous section.

The cluster variables are obtained through a principal component analysis. Six components are extracted, each of them contains a group of variables with the same spatial patterns (e.g. the component "employment" contains number of firms, number of stores, employment density and percentage of densely-built housing). Each sector receives, in a second step, a score on each of the six components (i.e. the higher the score, the better it matches with the component). Next, these standardized scores on the six components are the basis of a cluster analysis. Ward's Clustering Method is used to generate the necessary clusters. Nine clusters best reflected land use in the city region. The average factor scores on the components for each cluster indicate its spatial characteristics. For example, one cluster is the sector with a lot of highways, district roads and infrastructure and a low percentage of housing or facilities. Step by step the Spatial Zoning Algorithm (an example of one part of the algorithm is presented in Appendix A), which is based on the results of the cluster analysis, administrative and infrastructure borders, groups the sectors to larger zones, finally resulting in 33 destination choice zones (*see Figure 3)*. The 33 destination choice zones form the input for our model that aims at explaining the destination choice and the activity patterns on different spatial levels (*section 4*).

Figure 3: From 608 sectors to 33 Destination Choice Zones in the study area



Source: Van Hofstraeten and Verhetsel (2004)

4. Results

Before modeling destination choices, different combinations of variables are tested. The best result is obtained by the variables presented in Table 3 which summarizes and describes the data selected for modeling.

Table 3: Variables selected for modeling destination choices

Data source	Levels	Attributes	Types
OVG data	Individual	Age Gender	Continuous Binary (2 categories)
	Household	Income Location Household type	Discrete (5 categories) Discrete (4 categories) Discrete (8 categories)
	Trip	Purpose Mode of transport	Discrete (4 categories) Discrete (5 categories)
GIS	Zone	Land uses: - built-up housing - industrial / commercial /port area - agriculture and meadowland - housing and other developments;	Continuous Percentage of the geographical area by land uses type
		Density variables: - employment - shopping;	 Continuous Aggregated indicator by zone
	Trip	Travel time	Continuous Accessibility measure

The complexity of the data is obvious: different types of variables are combined for different data levels. However, by doing so, spatial planners are able to analyze sensitivities on different levels. The model can also explore socio-economic behavior and urban development simultaneously.

Let us first discuss the linearity of the utility function and the econometric models (Section 4.1). This will lead to the choice of the "best" model formulation. Modelling results and operational decisions are reported in Section 4.2 .

4.1. Impact of the utility function and the model choice on the modelling results

Table 4 compares the results obtained by a linear multinomial logit formulation to that of the Box-Cox estimation and the mixed logit with random coefficients. In the Box-Cox Logit, the estimated coefficient of variable λ is not significantly different from zero. In this case, the transformation of travel time is logaritmic (see equation [6]). For the model with random coefficients, the estimated variance of the distribution of the travel time coefficient is significantly different from zero; hence, we accept the hypothesis of a random coefficient of time: the perception of travel time varies randomly from one individual to another. These two alternative solutions to the standard MNL conduct to a drastic reduction in the log-likelihood: it drops from –4745 in the standard MNL to –3268 in de Box-Cox logit and to –2830 in the random coefficient mixed logit model. That is why the random coefficient formulation is preferred to the others.

Table 4 : Comparing results for different utility functions for travel time only.

	Multinomial Logit		Box-Cox Logit		Mixed Logit	
	Value	t-test	Value	t-test	Value	t-test
Accessibility variables						
Time	-0,90	-71,46	-4,97	-19,05	-	-
mean	-	-	-	-	-4,09	-28,17
variance	-	-	-	-	15,69	14,59
Lambda	-	-	0,01	0,32		
Number of parameters:	1		2		2	
Sample size:	3517		3517		3517	
Null log-likelihood:	-12297		-12297		-12297	
Final log-likelihood:	-4745		-3268		-2830	
Likelihood ratio test:	15104		18058		18934	
Rho-square:	0,61		0,73		0,77	

Table 5 : Parameter estimates for the multinomial logit and the nested logit

	Multinomial Logit		Nested Logit	
	Value	**t-test**	**Value**	**t-test**
Accessibility variables				
Time	-0,86	-67,99	-0,83	-63,11
Land Use variables				
Agriculture	0,034	7,75	0,027	6,20
Industry	-0,008	-2,10	-0,011	-3,09
Housing	0,021	14,60	0,019	13,44
Built up area	-0,012	-5,89	-0,011	-5,49
Size variables				
Employment	0,000021	2,49	0,000020	2,50
Number of shopping	0,002798	16,99	0,002422	14,59
Socio demographic variables				
Age	0,0155	3,14	0,0166	3,44
Income				
Income <500 euro	0,48	0,73	0,49	0,76
Household type				
Monoparental with 2 Children and more	1,13	1,97	1,16	2,04
Couple with 2 children and more	1,44	2,56	1,48	2,65
Household location				
suburb	-1,68	-5,71	-1,64	-5,70
Characteristic of chains				
Purpose				
Service	0,31	0,64	0,33	0,68
Mode				
bike	0,64	1,37	0,65	1,41
inclusive value 1			1	Fixed
inclusive value 2			1,24	28,71
inclusive value 3			1	Fixed
inclusive value 4			1,02	12,65
Number of estimated parameters:	14		16	
Sample size:	3517		3517	
Null log-likelihood:	-12297		-12297	
Final log-likelihood:	-4421		-4401	
Likelihood ratio test:	15753		15792	
Rho-square:	0,64		0,64	

Table 5 compares the modelling results obtained by the standard multinomial logit model against those obtained by the nested logit. In this first attempt, all explanatory variables were included in the models. The MNL is here considered as a benchmark. The log-likelihood slightly drops from –4421 in the MNL to –4401 on the nested logit and the coefficients of the inclusive factors for each branch of the nested logit are significantly different of 1. We hence can confirm that there is a hierarchical structure of stop choices in our studied area. In Antwerpen, an individual seems to first choose a broad zone (i.e. suburb, urban fringe, city centre, 19th century centre) and then within that zone, he chooses a more precise ward of destination. The nested logit enables one – in a certain way – to consider the nested effect of spatial scale as well as spatial correlation in destination choices. Let us also mention here, that as the coefficients of the inclusive variables are all significantly higher that 1, the IID does not anymore hold. The MNL is no more robust for explaining the destination choices. This gives us a good reason for preferring the nested logit formulation.

4.2. Modelling destination choice

In the previous section, the existence of a random coefficient in the travel time variable is shown. In Table 6, the results obtained in the estimation of the mixed multinomial logit are compared to those of the mixed logit. Compared to Table 5, we obtain a better fit in the quality on the estimators: the log-likelihood ratio is reduced and the Rho-Square is much higher. As shown in Table 5, the structure of choice is nested. The inclusive variables are significantly greater than 1. Given that the model includes a random coefficient, the mixed logit model is chosen as the best estimation of the spatial choice process in the case of Antwerpen (nested logit with random coefficient).

This confirms recent results obtained by Suarez *et al.* (2004) on the better fit of nested logit in selecting shopping centres. These authors also propose random effect models to perform the modelling. In this case, the structure of the error term can be decomposed as:

$$\varepsilon_{ji} = \mu_j + \omega_{ji} \qquad [11]$$

where μ_j represents the random geographical effect; it is assumed to be normally distributed. ω_{ji} is the residual effect which follows a Gumbel distribution. We end up with a mixed-logit model that takes into account the decomposition of the error term. Given that the performance of this kind of model is much lower than the previous one, it was not used in further analysis. We conclude that the most relevant model to explain the destination choice in the Antwerp city region is the mixed nested logit (random coefficients).

Table 6 : Parameter estimates for the multinomial logit and the mixed logit

	Mixed Multinomial Logit		Mixed Nested Logit	
	Value	**t-test**	**Value**	**t-test**
Accessibility variables				
Time : mean	-4,67	-23,54	-4,15	-22,86
Variance	20,21	12,36	16,21	11,79
Land Use variables				
Agriculture	0,030	4,15	0,029	4,63
Industry	0,028	5,00	0,025	4,95
Housing	0,038	15,80	0,031	12,80
Built up area	-0,023	-7,30	-0,017	-4,98
Size variables				
Employment	0,000019	2,01	0,000006	0,56
Number of shopping	0,004842	17,29	0,004320	14,13
Socio demographic variables				
Age	0,0003	0,06	0,0007	0,09
Income				
Income <500 euro	0,33	0,57	0,33	0,36
Household type				
Monoparental with 2 Children and more	1,03	1,94	1,03	1,12
Couple with 2 children and more	1,19	2,39	1,20	1,52
Household location				
suburb	-1,70	-4,88	-1,62	-3,41
Characteristic of chains				
Purpose				
Service	0,14	0,28	0,16	0,21
Mode				
bike	0,37	0,89	0,39	0,58
inclusive value 1			1	Fixed
inclusive value 2			1,85	3,55
inclusive value 3			1	Fixed
inclusive value 4			2,64	2,46
Number of estimated parameters:	15		17	
Sample size:	3517		3517	
Null log-likelihood:	-12297		-12297	
Final log-likelihood:	-2582		-2552	
Likelihood ratio test:	19430		19490	
Rho-square:	0,79		0,79	

The travel time variable, i.e. an accessibility measure, has a high significance compared to other variables of the model. This means that travel time between origin and destination is able to explain a large part of the model's variability. Destinations located further away are less interesting than those closer to the home address.

The choice of a destination is additionally influenced by spatial variables and socio-economic variables:

- Spatial variables include density and land use information. The influence on the number of shopping and employment alternatives is positive: they increase the attractiveness of destination. The higher the percentage of housing and industrial surface in the zone, the higher the probability to choose the alternative. In contrast, the land use variables built-up area has a negative effect on the attractiveness of the destination. These results correspond to our expectations.

- Socio-economic variables include age, income and household type. These variables affect all the destinations simultaneously: e.g. increasing the mean age of the population in an origin zone affects the probability of all alternatives in the choice set, including the origin zone. In other words, besides having an impact in one specific zone, the socio-economic variables also explain the flux between the origin (individual location) and the destination.

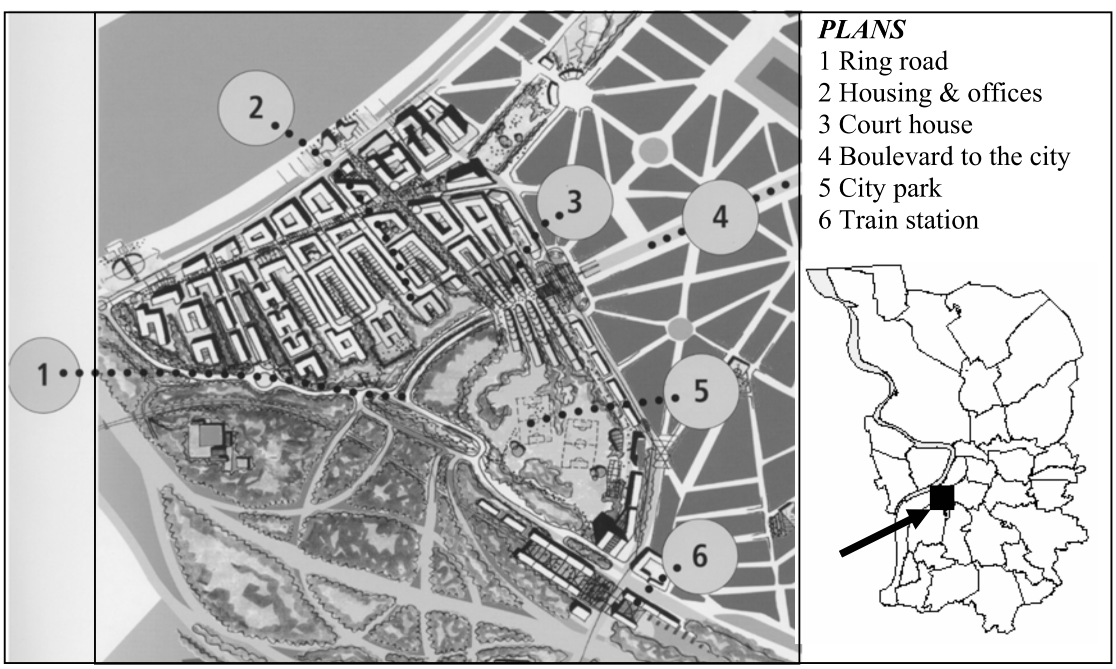
4.3. Simulations

Besides analyzing the MNL results, it is also interesting to analyze the impact of a specific change in variables on the choice of an alternative. To illustrate the simulation process, it is now explained what would happen if two urban development projects are introduced in the Antwerpen city region, the so-called Nieuw-Zuid and Petroleum-Zuid.

Nieuw-Zuid is a project that aims at connecting the Southern suburbs with the Antwerpen city centre. This prestigious project contains a new Court, new residential areas along with office space, a public park and retail space. It is located near the Scheldt River and nowadays the area is characterized by waste land and obsolete port infrastructure. In order to make these new development areas more easily accessible by public transport, plans also contain the expansion of the neighboring railway station and the development of extra tram lines to this area (Stad Antwerpen, 2002). Figure 4 maps the Nieuw-Zuid project.

In this simulation, the calculated spatial variables on land use, density and accessibility for the statistical sector (which encloses the urban development project Nieuw-Zuid) are modified according to the architect's plans. Nieuw-Zuid will generate 13 hectares of housing area (equal to approximately 3000 new inhabitants), 8 hectares of offices (equal to 200 000 square meters of floor-space) and 9 hectares of municipal park (Vespa, 2004). Appendix B shows the estimated changes in spatial variables.

Figure 4: Plan of the urban development project Nieuw-Zuid



Source: Stad Antwerpen (2002)

On the other side of the ring road, another urban development project is planned in the near future: **Petroleum-Zuid**. The urban development area of Petroleum-Zuid is still part of the port area, even if most port activities left the site years ago. One of the major goals is to open the southern border of the city of Antwerpen and connect some less prosperous neighborhoods with the inner city. Today only five companies are still active in this zone. Land destination plans at regional level point out the area as industry and park zone. However, it is not clear yet what exactly is understood by park zone. More specifically, plans are to redevelop the urban area as a zone of mixed-use, namely as a city project (with economic activities) in green setting (Van Dyck, 2003). A study now examines the impact of the development of an industrial area of 75 to 80 hectares. It is proposed to create more public space (either a municipal park or a square) on a total area of 20 hectares (VESPA, 2004).

Both urban development projects are situated in the same destination choice zone. By this, next to calculating the impact separately of one project on the attractiveness of its sector, it is also possible to study the effect of introducing both projects simultaneously on the probability to choose the destination choice zone to which both sectors are assigned. These results presented in Table 11.

Table 11: Impact on the urban development projects

	Sector		**Destination zone**
	Nieuw Zuid	*Petroleum Zuid*	*(including Nieuw Zuid and Petroleum Zuid)*
Change in attractiveness	+78.2%	-65.2%	+12.7%

According to our simulation results, the Nieuw Zuid project will be more attractive to inhabitants of the Antwerpen city region than the Petroleum Zuid project. This can be explained by the fact that a project containing new shopping facilities, a municipal park, housing and offices obviously attracts more people than an industrial project containing no residential or shopping facilities. Finally, the positive effect on the probability to choose the destination zone of both projects can be noticed in Table 11.

5. Conclusions

This paper deals with the introduction of the characteristics of space in destination choice modeling. Several modeling alternatives are suggested and compared. The application is limited to the city region of Antwerp (Belgium) and to one data set (OVG). Results seems to be quite promising methodologically and empirically.

When considering space in destination modeling, the main methodological problems encountered are: (1) summarizing the spatial reality by a few variables, (2) defining independent spatial alternatives, (3) the availability of geographical information systems (adequate data, software and "lifeware"), and (4) choosing an adequate formulation for the model choice. This explains why only a few approaches of this problem are to be found in the former discrete choice literature. In our case, several variables were created in order to "measure" space and spatial attractiveness. Several modeling methodological formulations were also developed and compared in order to avoid the numerous methodological pitfalls of discrete choice modeling; in our case, the mixed nested logit (MXNL) seems the be the best formulation.

The application consists in comparing different formulations of the model, interpreting the parameters for the present situation in Antwerp but also considering new infrastructures in the city planning process. By using this approach, public stake holders can encourage developments in specific areas and study the impact of this policy measure on the overall mobility. They can also play an active role in investing in land development, housing, infrastructure, etc.

Further analyses will be done. Besides simulating techniques, the calculation of elasticities, which specify the proportional demand increase or decrease caused by a one-percent change in variable, can also be calculated. The elasticities will be computed for spatial and socioeconomic variables. Further simulations will enable to test the sensitivity of the model to changes in spatial and behavioral measures and statistical modeling choices. Avenues for future research consist in further developing the model and in adding new variables in the explanatory process. Furthermore, the techniques developed in this paper for generating spatial variables, defining destination choice zones and destination choice modeling will be applied to other city regions (e.g. Gent, Mechelen.) in order to compare the empirical results and further test the robustness of the methods.

Acknowledgements

The research is performed in execution of the Second Scientific Support Plan for a Sustainable Development Policy (SPSD II) for the account of the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs (OSTC). We thank the Flemish Government, in particular Eddy Peetermans for the provided practical assistance with the GIS applications and Wilfried Goossens for providing OVG data. The authors are also grateful to the public transport company “De Lijn” for providing useful data deliverance.

References

- Axhausen, K. & Gärling, T. (1992) “Activity-Based Approaches to Travel Analysis: Conceptual Frameworks, Models and Research Problems” *Transport Reviews*, Vol. 12(4), pp. 323-341.
- Bates, J. (2000) “History of Demand Modeling”, In: D.A. Hensher & K.J. Button (Eds.), *Handbook of Transport Modeling*, Pergamon, Oxford, pp. 11-33.
- Badoe, D.A. & Miller, E.J. (2000) “Transportation-Land-Use Interaction: Empirical Findings in North America and their Implications for Modelling” *Transportation Research D*, Vol. 5, pp. 235-263.
- Ben-Akiva, M. & Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge (Mass.) and London, MIT Press.
- Bhat, C. & Zhao H. (2002) “The Spatial Analysis of Activity Stop Generation” *Transportation Research B,* Vol. 36(7), pp. 593-616.
- Cervero, R. (1996) “Mixed Land-Use and Commuting: Evidence from the American Housing Survey” *Transportation Research A*, Vol. 30(5), pp. 361-377.

- Cervero, R. & Knockelman, K. (1997) “Travel Demand and the 3D’s: Density, Diversity and Design” *Transportation Research D*, Vol. 2(3), pp. 199-219.
- Cirillo, C., Cornélis, E., Legrain, L. & Toint, P. (2003) “Combining Spatial and Temporal Dimensions in Destination Choice Models” Transportation Research Group, FUNDP, Namur. Unpublished.
- Cramer, J.S. (1991), *The Logit Model: An Introduction for Economists*, London, Edward Arnold.
- Dijst, M. & Vidakovic, V. (1997) “Individual Action-Space in the City”, In: D.F. Ettema & H.J.P Timmermans (eds), *Activity-Based Approaches to Travel Analysis*, Pergamon, Oxford, pp.117-134.
- Domencich, T. & McFadden, D. (1975) *Urban Travel Demand: A Behavioural Analysis*, North-Holland Publishing Company, Amsterdam, The Netherlands.
- ESRI (2003), *Technical Article, No. 18856*. Downloadable from website: http://support.esri.com
- Hensher D.A. & Greene, W. (2002) “Specification and Estimation of the Nested Logit Model: Alternative Normalisations” *Transportation Research B*, 36, 1-17.
- Hensher, D.A. & Johnson, L.W. (1981), *Applied Discrete-Choice Modeling*, Croon Helm, London.
- Johnson, N.L. & Kotz, S. (1970) *Continuous Univariate Distributions-1*, New York, Houghton Mifflin Company.
- Jones, P.M., F. Koppelman, & J.P. Orfeuil (1990) “Activity Analysis: State-of-the-art and Future Decisions”, In: P.M. Jones (Ed.) *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Gower, Aldershot, UK
- Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.
- McFadden, D. (1973) “Conditional Logit Analysis of Qualitative Choice Behavior” In: P. Zarembka (Ed.) *Frontiers in Econometrics*, Academic Press, New York.
- McFadden, D. (1976) “Quantal Choice Analysis: A Survey”, *Annals of Economic and Social Measurement*, Vol. 5(4), pp.363-390.
- McFadden, D. (1978) “Modelling the Choice of Residential Location”, In: A. Karlqvist, L. Lundqvist, F. Snickars & J.W. Weibull (Eds.) *Spatial Interaction Theory and Planning Models*, Amsterdam, North-Holland Publishing Company, pp. 75-96.
- McNally, M. (2000) “The Activity-Based Approach”, In: D.A. Hensher & K.J. Button (Eds.), *Handbook of Transport Modeling*, Pergamon, Oxford, pp. 53-69.
- Miller, H.J. (1999) “Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T)” *Geographical Analysis*, Vol. 31, pp. 373-399.
- NIS, (2001) *Bevolking van het Rijksregister op 1 januari 2001*, Databank NIS
- OC-GIS Vlaanderen (2002), *Bodembedekkings- en Bodemgebruiksbestand*, CD-ROM.
- Openshaw, S. & Taylor, P.J. (1979) "A Million or So Correlated Coefficients: Three Experiments on the Modifiable Areal Unit Problem," In: N. Wrigley & R.J. Bennet (Eds.), *Statistical Applications in the Spatial Sciences*, Pion, London.
- Ortuzar, J.D. & Willumsen, L.G. (2001) *Modeling Transport*, John Wiley & Sons, Chichester.

- Papola, A. (2004) ”Some Developments on the Cross-Nested Logit Model” *Transportation Research B*. Article in press.
- Revelt, D., & Train, K. (1999) “Customer-Specific Taste Parameters and Mixed Logit,” working paper, Department of Economics, University of California, Berkeley. Paper available on: http://elsa.berkeley.edu/users/train/index.html
- Sherlock, R., Mooney, P., Winstanley, A. & Husdal, J. (2002) *Shortest Path Computation: A Comparative Analysis*, Paper presented at the GISRUK 2002 Conference, University of Sheffield, Sheffield.
- Stad Antwerpen (2002), *Nieuw Zuid: De Nieuwe Place To Be,* Antwerpen. Unpublished.
- Stead, D. (2001) “Relationships between Land Use, Socio-Economic Factors and Travel Patterns in Britain” *Environment and Planning B*, Vol. 28(4), pp. 499-529.
- Suarez A., Rodriguez del Bosque I., Rodrigues-Poo J. & Moral I. (2004) “Accounting for heterogeneity in shopping, centre choice models” *Retailing and Consumer Services*, Vol.11, pp. 119-129.
- Timmermans, H.J.P. & Golledge, R.G. (1990) “Applications of Behavioural Research on Spatial Problems II: Preference and Choice”, *Progress in Human Geography*, Vol. 14(3), pp. 311-354.
- Train, K. (1986) *Qualitative Choice Analysis, Theory, Econometrics, and an Application to Automobile Demand*, Cambridge (Mass.) and London, MIT Press.
- Train K. & MacFadden D. (2000) “Mixed MNL Models for Discrete Response” *Journal of Applied Econometrics*, Vol.15, pp. 447-470.
- Van Dyck, B. (2003) *Brownfields Redevelopment: Governance and Partnerships*, Erasmus University Rotterdom/EURICUR, The Netherlands.
- Van Hofstraeten, D. & Verhetsel, A. (2003) *De Introductie van Ruimtelijke Variabelen en Zoneringstechnieken in de Analyse van het Verplaatsingsgedrag*, Unpublished.
- Van Wee, B. (2002), ‘Land Use and Transport: Research and Policy Challenges’, *Journal of Transport Geography*, Vol. 10, pp. 259-271.
- Verhetsel, A., Witlox, F., Tindemans, H. & Van Hofstraeten, D. (2004), *SAMBA: Spatial Analysis and Modeling based on Activities. A pilot study for Antwerp and Ghent (Flanders, Belgium)*, In: K. Williams (Ed.) *Travel/Mobility and Urban Form,* Ashgate Publishers (forthcoming).
- Wise, S. (2002) *GIS Basics*, Taylor & Francis, London.

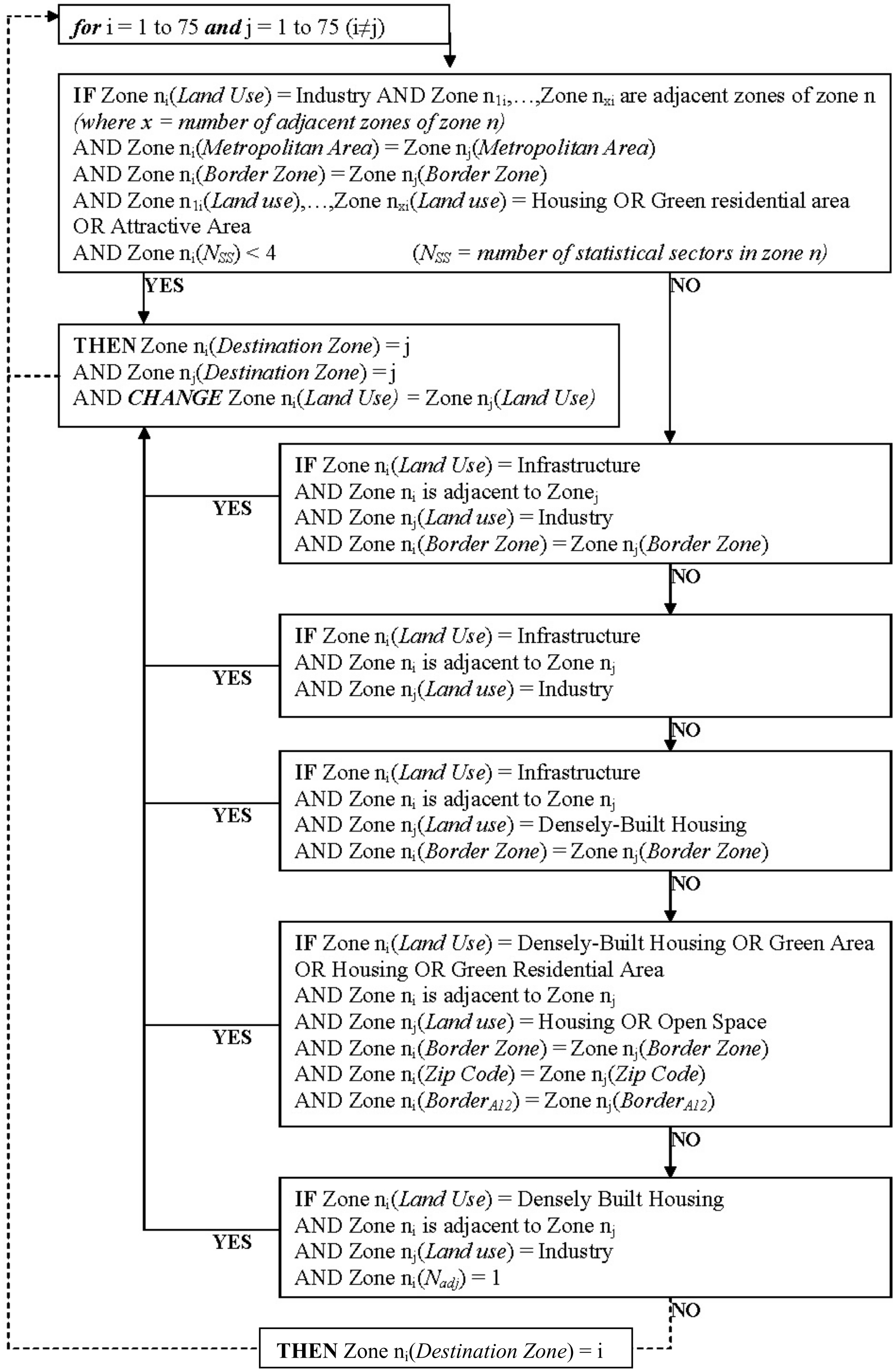
Websites

- SCOOT, Internet search engine: http://www.scoot.be

Other Sources

- VESPA (Isabelle Van Achter, Koen Der Kinderen)

Appendix A: Example of the spatial zoning algorithm



Appendix B: Impact of the urban development project Nieuw-Zuid

	LAND USE 2001	CHANGE	ESTIMATED LAND USE
Housing & development	0,19%	+33,63%	33,82%
Municipal Parks	0,00%	+13,68%	13,68%
Highway roads	7,45%	+2,63	10,08%
District roads	0,00%	+2,57%	2,57%
Alluvial Meadowland	0,01%	+0,01%	0,00%
Mixed Forests	0,41%	-0,18%	0,23%
Broad-Leaved Forests	13,05%	-1,30%	11,75%
Meadowland	7,59%	-6,47%	1,12%
Infrastructure	13,60%	-6,69%	6,91%
Port Infrastructure	28,69%	-18,80%	9,89%
Agriculture/open space	25,82%	-19,69%	6,13%
Densely Built Housing	0,24%	0	0,24%
Water	2,67%	0	2,67%
Land use total	*99,71%*		*99,04%*
	DENSITY	**CHANGE**	**ESTIMATED DENSITY**
Population	1	+2999	3000
Population density	1,02 / km²	+2455,33 / km²	2456,35 / km²
Employment	7	+1993	2000
Employment density	7,12 / km²	+1957,96 / km²	1965,08 / km²
Number of firms	1	+49	50
Number of shops	0	+40	40
Shopping density	0 / km²	+39,30 / km²	39,30 / km²
Number of schools	0	0	0
School density	0 / km²	0	0 / km²
	ACCESSIBILITY	**CHANGE**	**ESTIMATED ACCESSIBILITY**
Accessibility to shopping	0,61	+0,59	1,20
Accessibility to industry	0,13	+0,07	0,20
Accessibility to housing	1,27	+0,13	1,40
Public transport freq/hour	44,19	+500	544,19

Appendix C: Impact of the urban development project Petroleum-Zuid

	LAND USE 2001	CHANGE	ESTIMATED LAND USE
Industry	1,25%	+44,62%	45,87%
Municipal Parks	0%	+10%	10%
Alluvial Meadowland	0,63%	+0,19%	0,82%
Housing & development	0%	0	0%
Highway roads	0,46%	0	0,46%
District roads	0%	0	0%
Densely Built Housing	0%	0	0%
Water	4,23%	0	4,23%
Mixed Forests	0,04%	-0,04%	0%
Broad-Leaved Forests	10,92%	-2,53%	8,39%
Meadowland	11,51%	-6,41%	5,10%
Agriculture/open space	15,18%	-8,71%	7,01%
Infrastructure	16,91%	-11,13%	5,78%
Port Infrastructure	33,18%	-20,40%	12,78%
Land use total	*100%*		*100%*
	DENSITY	**CHANGE**	**ESTIMATED DENSITY**
Population	15	0	15
Population density	9,96 / km²	0	9,96 / km²
Employment	533	+521	1054
Employment density	354 / km²	+346 / km²	700 / km²
Number of firms	10	+20	30
Number of shops	0	0	0
Shopping density	0 / km²	0	0 / km²
Number of schools	0	0	0
School density	0 / km²	0	0 / km²
	ACCESSIBILITY	**CHANGE**	**ESTIMATED ACCESSIBILITY**
Accessibility to shopping	0,53	+0,08	0,61
Accessibility to industry	0,12	+0,30	0,42
Accessibility to housing	1,14	+0,13	1,27
Public transport freq/hour	0	0	0